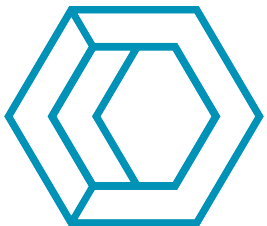


Hoax Detection Mechanism in the Social Media

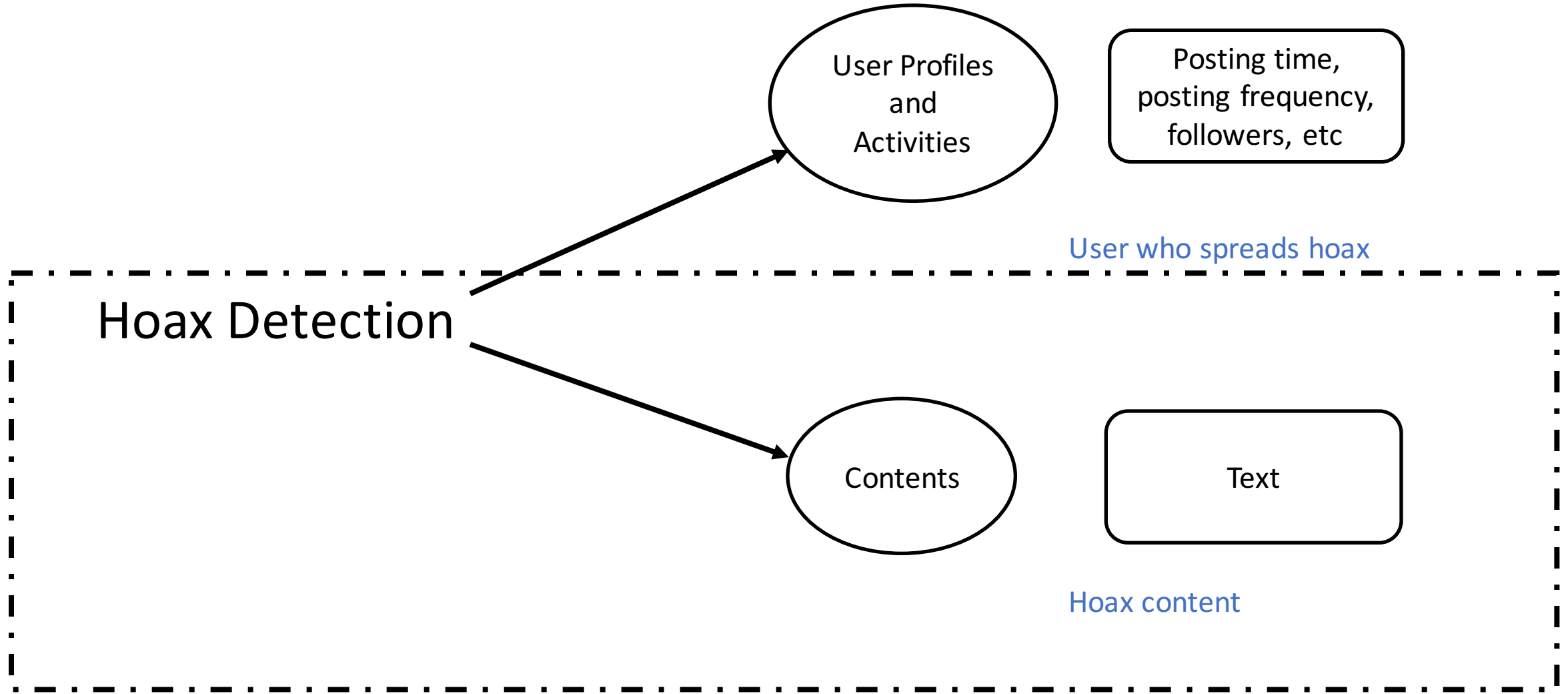
Canggih Puspo Wibowo

24-25 April 2019

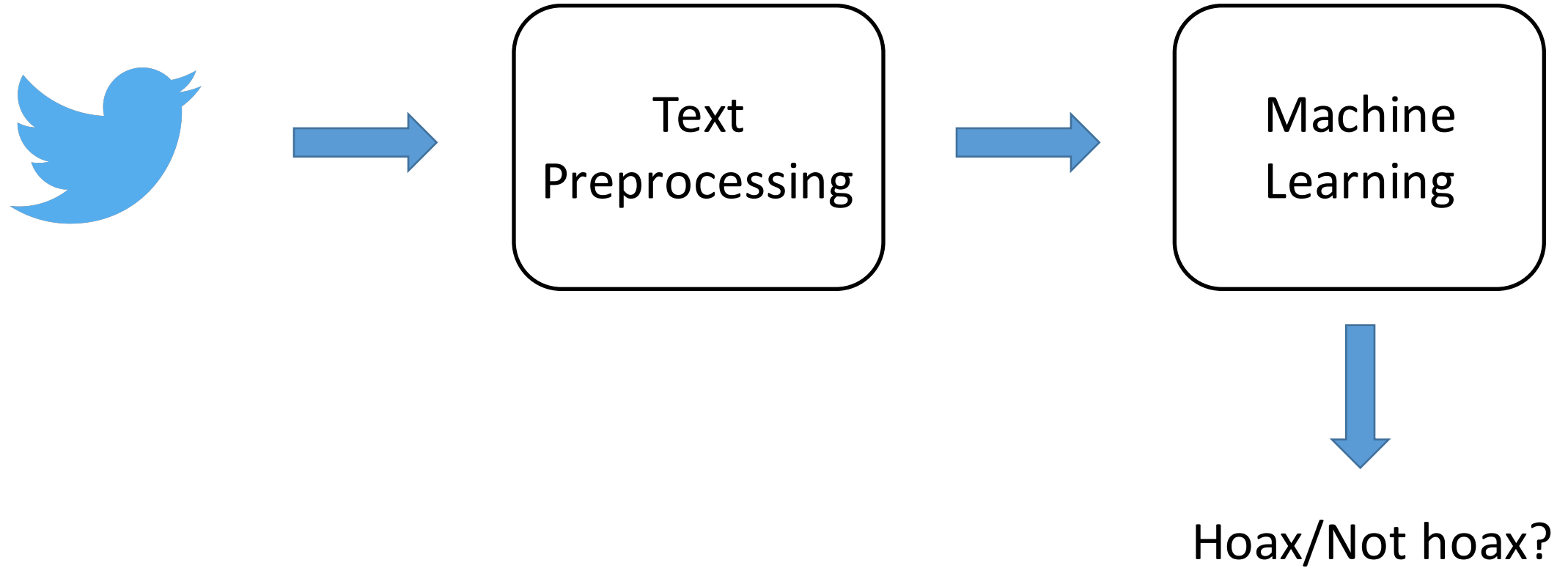


KEDATA

Hoax Detection Types



Hoax Detection Mechanism



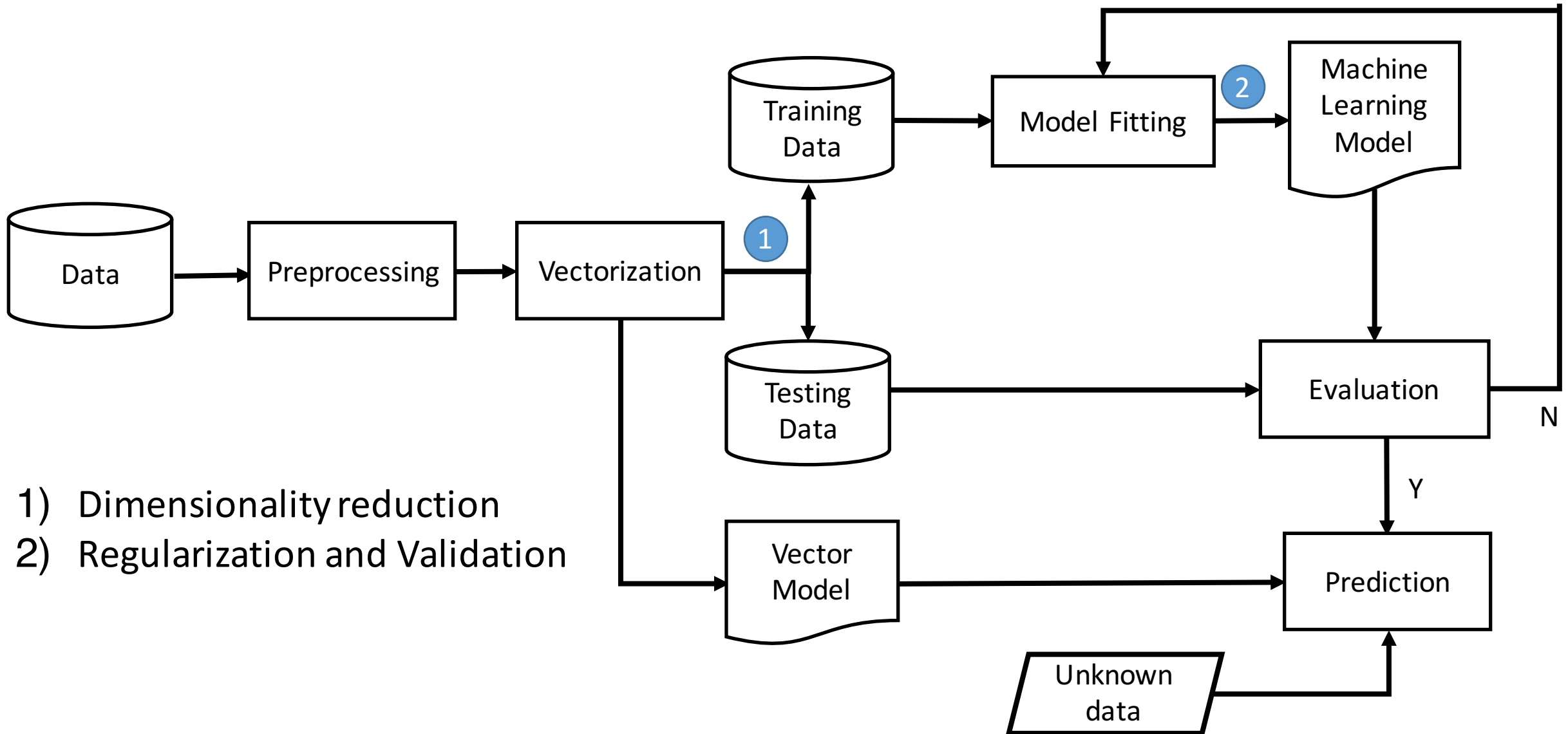


Hoax Detection System

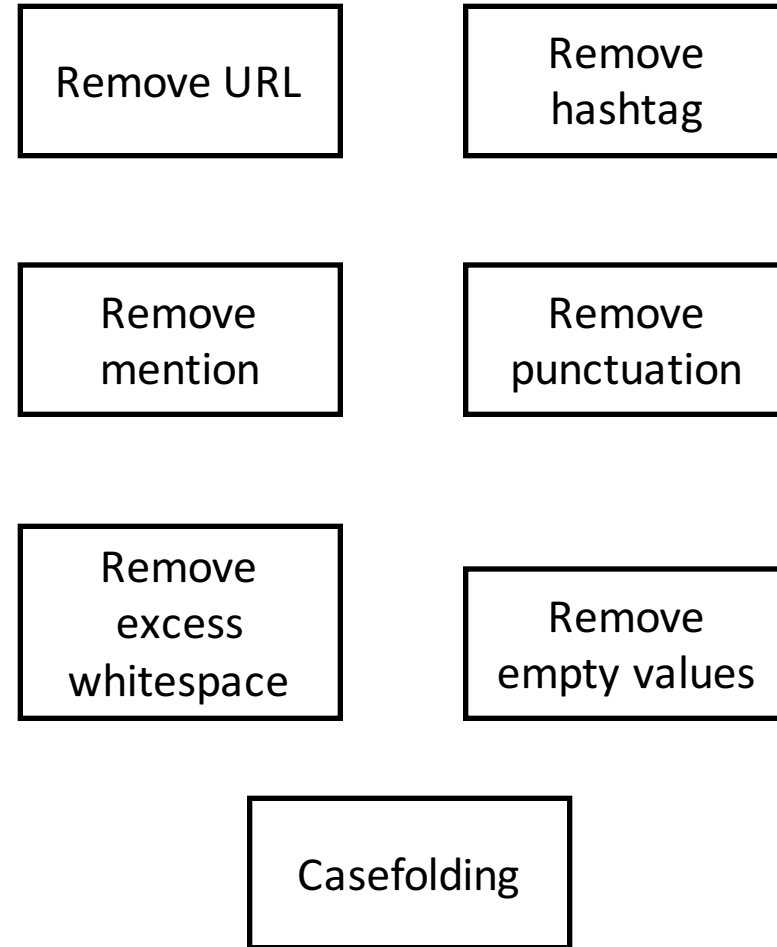
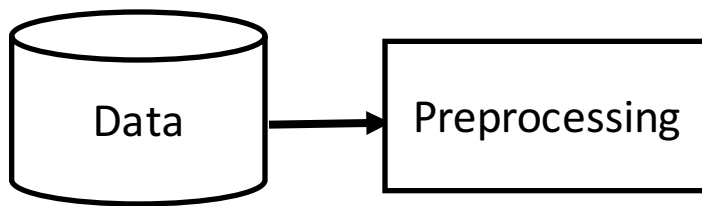
Canggih Puspo Wibowo



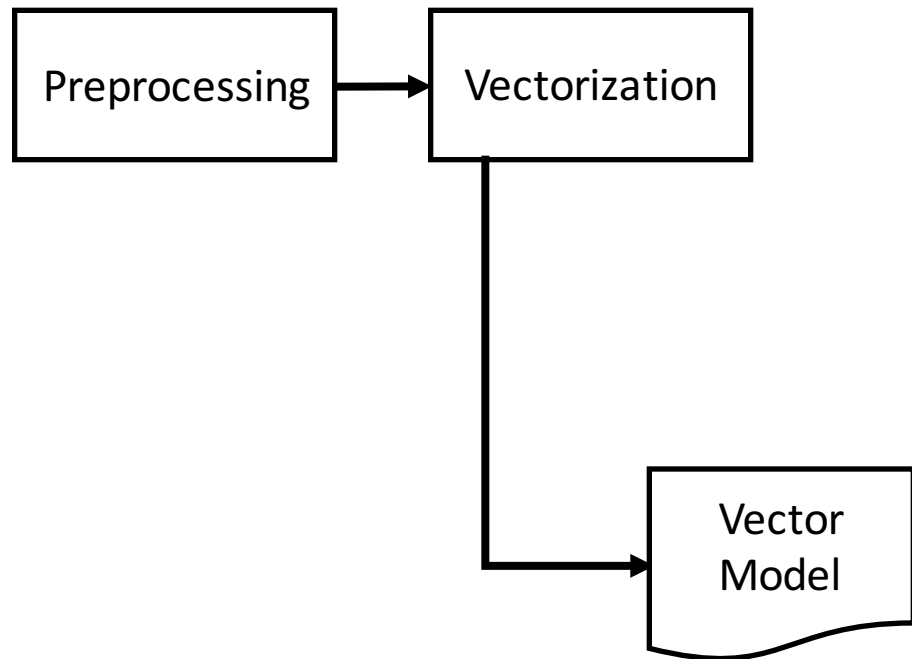
Simplified Hoax Detection Flowchart



Preprocessing



Vectorization



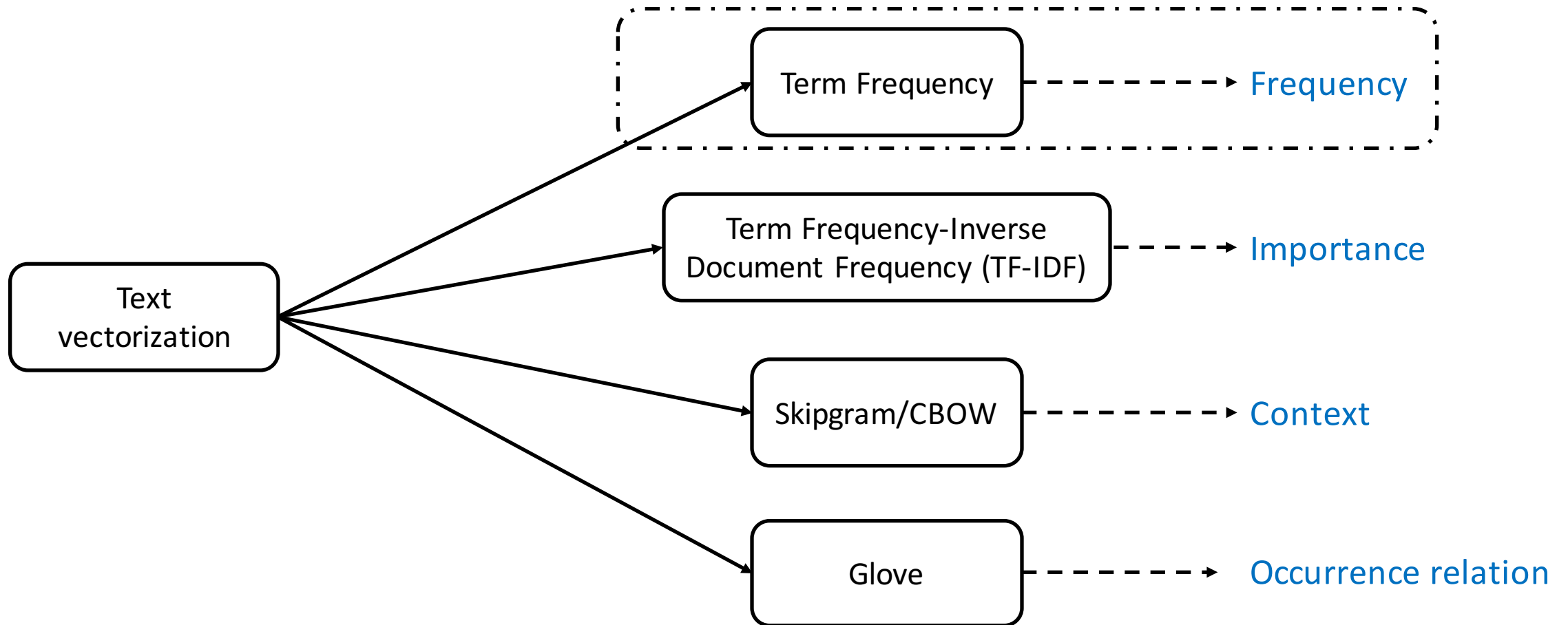
text to numeric

Ex:

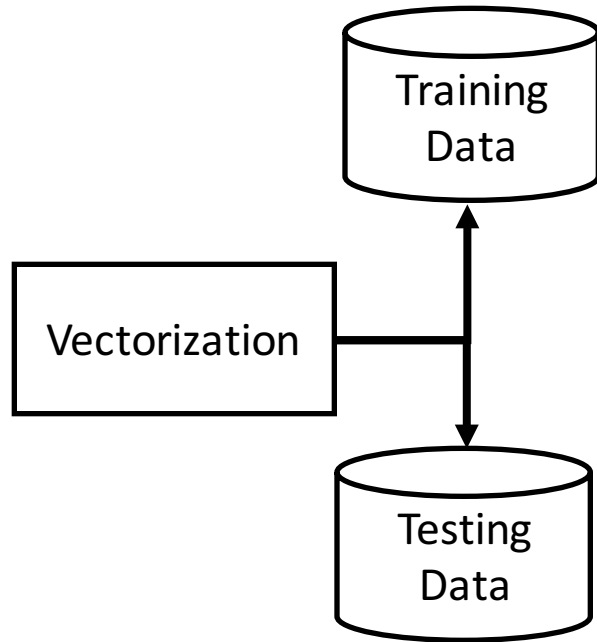
“saya” → [4,1,5,6,7,3,5,6,3,...]

“saya makan nasi” → [2,1,4,2,1,1,...]

Vectorization: Methods



Split into Training and Testing



What is the ratio?

Common rule (pareto):

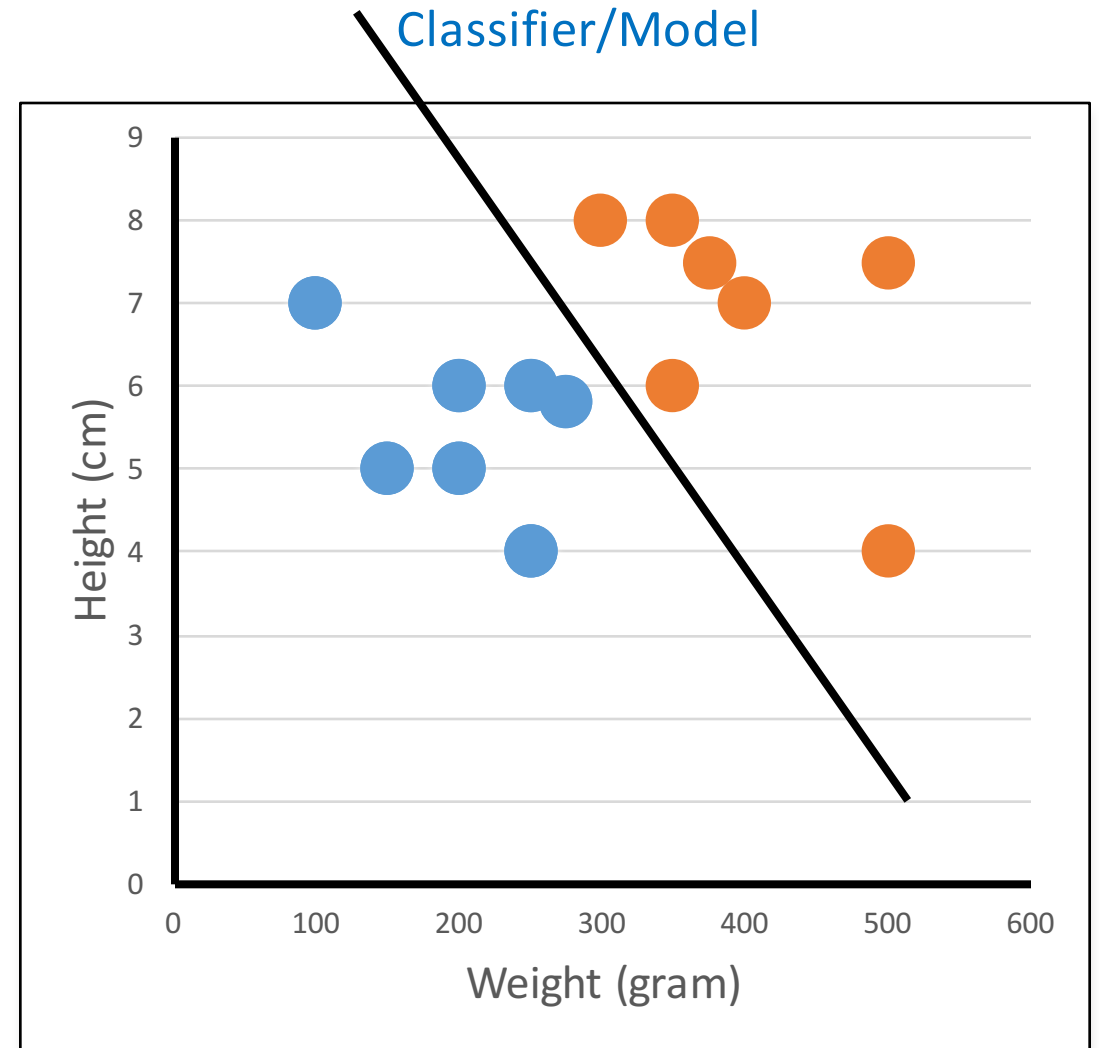
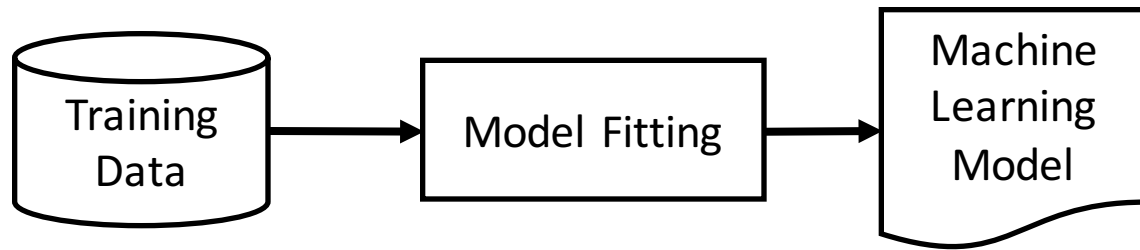
Training = 80%

Testing = 20%

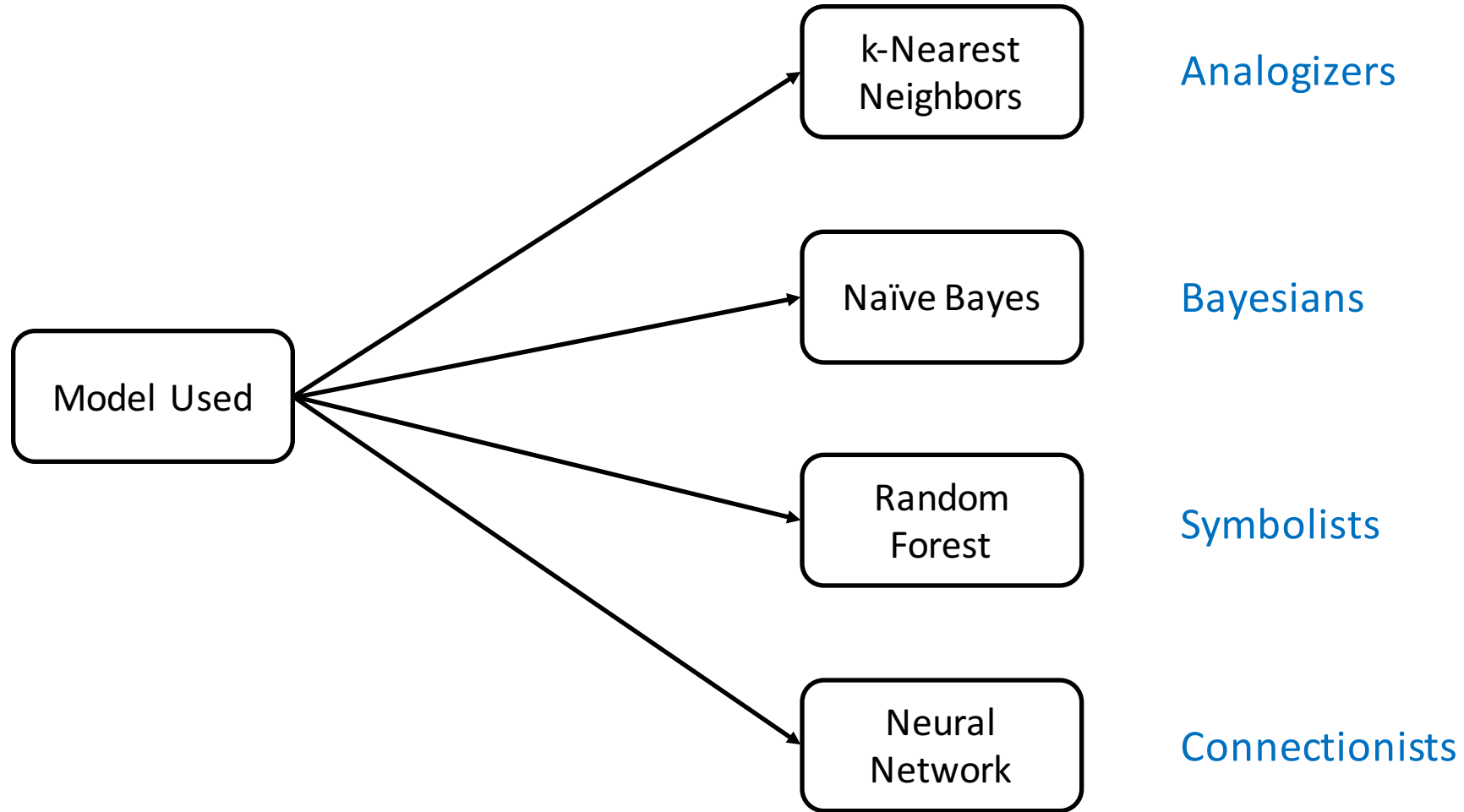
Factors concerning ratio:

- 1) Variance of data
- 2) Computation

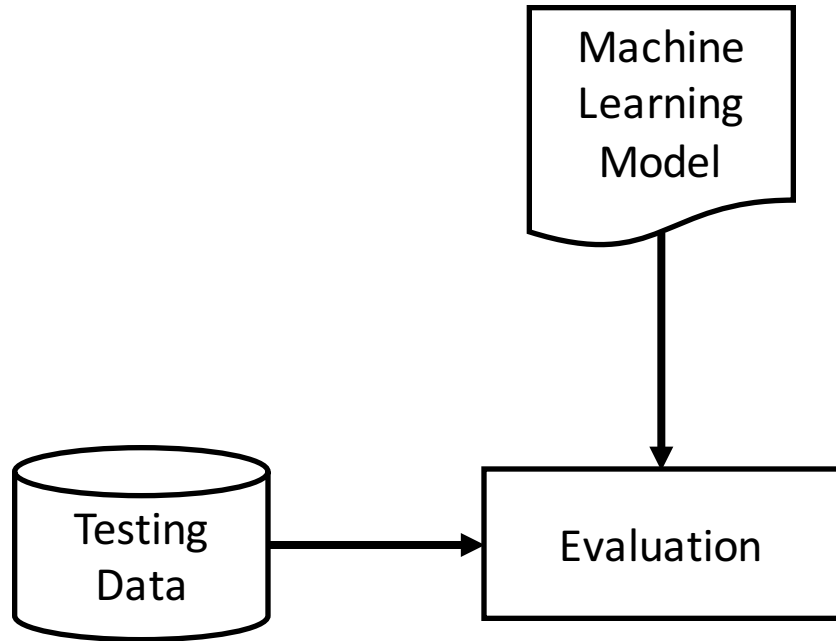
Model Fitting



Model Fitting (2)



Evaluation



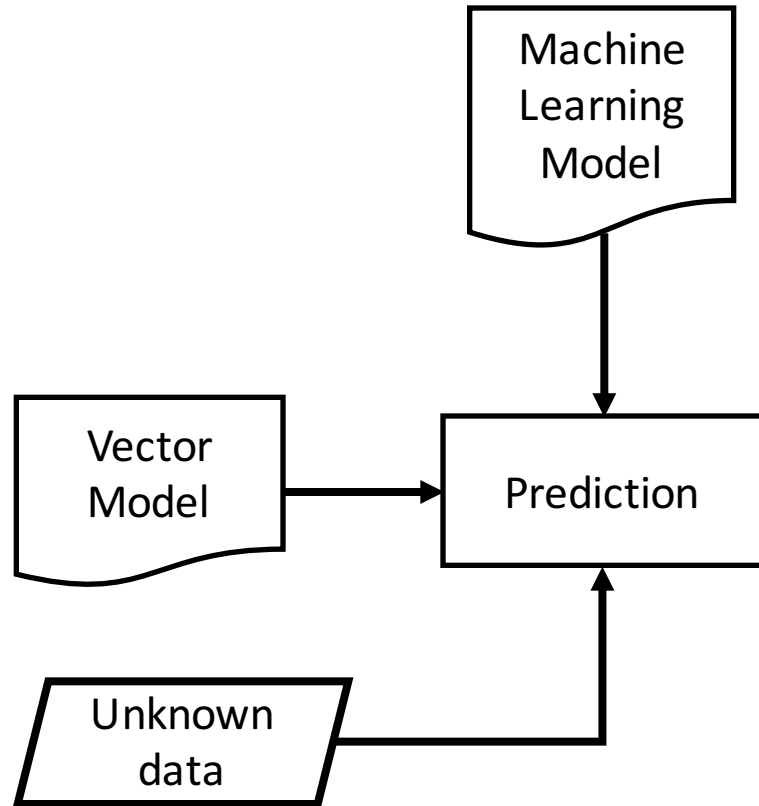
H0 : hoax

H1 : not hoax

		Actual	
		Hoax	Not hoax
Prediction	Hoax	TP	FP
	Not hoax	FN	TN

Accuracy, Precision, Recall, F1 score

Prediction



“predict unknown data”

“This cannot be hoax!!”

Preprocessing

“this cannot be hoax”

Vector Model

[4,5,1,3,6,0,0,...]

Machine Learning Model

95% hoax